IMPORTANT OMITTED SPATIAL VARIABLES IN SAFETY MODELS: UNDERSTANDING CONTRIBUTING CRASH CAUSES AT INTERSECTIONS

Sudeshna Mitra, California Polytechnic State University, Civil & Environmental Engineering Department 13-263, San Luis Obispo, CA 93407-0353, Tel. (805) 756-2947, Fax. (805) 756-6330 sudeshna@email.arizona.edu

Simon Washington, Arizona State University, Department of Civil & Environmental Engineering, P. O. Box 875306, Tempe, AZ 85287-5306, Tel. (480) 965-2220, Fax. (480) 965-0557 simonw@engr.arizona.edu

Ida van Schalkwyk, Arizona State University, Department of Civil & Environmental Engineering, P. O. Box 875306, Tempe, AZ 85287-5306, Tel. (480) 965-3001, Fax. (480) 965-0557 idavan@asu.edu

ABSTRACT

Advances in safety research—trying to improve the collective understanding of motor vehicle crash causation—rest upon the pursuit of numerous lines of inquiry. The research community has focused much on analytical methods development (negative binomial specifications, simultaneous equations, etc.), on better experimental designs (before-after studies, comparison sites, etc.), on improving exposure measures, and on model specification improvements (additive terms, non-linear relations, etc.).

One might think of different lines of inquiry in terms of 'low lying fruit'—areas of inquiry that might provide significant improvements in understanding crash causation and/or prediction. It is the contention of this research that omitted variable bias caused by the exclusion of important (causal or surrogates of causal) variables is an important line of inquiry in safety research. In particular, spatially related variables are often difficult to collect and omitted from crash models—but offer significant ability to better understand contributing factors to crashes.

This study—believed to represent a unique contribution to the safety literature—develops and examines the role of a sizeable set of spatial variables in intersection crash occurrence. In addition to commonly considered traffic and geometric variables, examined spatial factors include local influences of weather, sun glare, proximity to drinking establishments, and proximity to schools—representing a mix of potential environmental and human factors influences. The results indicate that inclusion of these factors results in significant improvement in model explanatory power, and the results also generally agree with theoretical expectation. The research illuminates the importance of environmental and human factors spatial variables in safety research and also the negative consequences of their omission.

INTRODUCTION

Safety research spans the range of inquiries including analytical method improvements (negative binomial specifications, simultaneous equations, etc.), improved experimental designs (before-after studies, comparison sites, etc.), assessing alternative exposure metrics, and model specification improvements (additive terms, non-linear relations, etc.). Relatively little research has focused on the inclusion of traditionally excluded or omitted variables. In particular, variables that are related to spatial factors that are typically unavailable in crash databases have not been examined in great detail. There are numerous factors that might be used to improve our understanding of crash occurrence at intersections.

Intersections are obvious points on the transportation system where omitted variables can improve the understanding of crash contributors. At-grade intersections are the most complex locations in the transportation system with relatively large numbers of conflicts and collisions. Not surprisingly, the safety of at-grade intersections is an important concern for road traffic engineers. Safety performance functions (SPFs) are often developed and used to predict safety (crashes) as a function of exposure (traffic volumes). There have been considerable analytical improvements in the past few decades focusing on the appropriate specification of SPF, including variables selection and functional forms. For example, linear regression models were deemed inappropriate for modeling SPFs and researchers like Joshua and Garber (1), Miaou and Lum (2), and Miaou (3) suggested that count models should replace them. Researchers have also evaluated and debated the use of modified count variables, such as zero-inflated or hurdle models (4-7) panel data models (8-9), and so on. There are also a considerable number of studies (10-14) focusing on the appropriate functional form of the count regression models.

While accurate model predictions rely on the choice of correct mathematical relationship between variables and correct distributional assumptions, the selection of a comprehensive and 'correct' set of independent variables is arguable *more* important. Omission of important variables in the crash process will introduce bias in model estimation. The analysis of omitted variable bias is an important research area in econometric research; however, in transportation safety research, omitted variable problems are relatively unexplored. The primary reason for its scarcity is the practical constraints on data availability, which severely limit the number and type of variables that can be included in crash models. Moreover, the inclusion of traffic volume in correct form and as the primary predictor of crash occurrence has been shown to explain the majority of variance in crashes—thus rendering additional variables as second-order effects—effects of relatively lesser importance. While traffic volume (exposure) is indisputably the most important factor, other geometric, environmental, and spatial factors may play an important and complex role in crash occurrence. The safety effect of many of these factors has been examined in prior research, such as the effects of various geometric designs, weather effects such as rainfall, snow etc., and some human related factors. However, the effects of numerous spatial factors have received relatively sparse attention.

There are many environmental and human factors spatial factors that on theoretical grounds should affect and the frequency and/or severity of motor vehicle crashes. Some of these factors have been examined extensively while others have not. These spatially related factors include:

- Urban versus rural locations experience different safety effects. Although this urban-rural distinction is a blurry one—and merely serves as an indicator for 'other' differences—it is often used to differentiate safety. The main differences underlying urban and rural areas are the driving population, the intensity of traffic volumes, the magnitudes of the peak periods, roadside distractions, and the complexity of the driving environments.

- The location of special traffic generators such as schools, colleges and universities, or stadiums may impact safety. Neighborhoods with schools, especially elementary and middle schools may be associated with greater numbers of pedestrian and bicycle crashes. Colleges and universities may also result in higher number of pedestrian, bicycle, and even drunk-driving related crashes. Stadiums that serve alcohol might be associated with higher numbers of drunk-driving crashes on days of sporting events.

- Socio-demographics also tend to be spatially related. Age tends to be a reasonable predictor of such things as risk-aversity, driving experience, and physical abilities (e.g. perception and reaction times, vision, and hearing ability). The National Highway Safety Administration (NHTSA) traffic safety facts repeatedly indicate that the young persons between 16 and 20 years old have the highest fatality and injury rates per 100,000 population. On the other hand older drivers are more involved in crashes per mile driven. The distribution of these types of drivers is rarely randomly distributed throughout a network, with neighborhoods, communities, or regions often under- or over-represented by certain socio-demographics.

- The location of drinking establishments—bars with the primary intent to sell and serve alcohol in particular—often occur in clusters within a region and represent locations where intoxicated drivers are more likely. Each year about 40% of fatal crashes are alcohol related. In 2004, 9% of the injured persons received their injuries in alcohol-related crashes. The proximity of transportation network locations (e.g. intersections, freeways, etc.) to these drinking establishments may prove to be important predictors of DUI-related crashes.

- Weather effects can also be spatial. Rain, snow, and ice may not be evenly distributed throughout the transportation network. Sun glare can be detrimental at certain locations during certain times of day and times of year. Icing may occur at certain locations under specific weather conditions. Of course these effects often compromise safety and expose drivers to higher levels of risk.

- Police enforcement campaigns might be spatially related to where moving violations occur or where traffic is less constrained. Policies adopted by local and state law enforcement may influence safety when enforcement is increased in certain corridors or regions. Safety, as a result, may be improved in these locations.

Some of these spatial factors have been previously examined by researchers. For example, Ivan et al. (15) and Ossenbruggen et al. (16) examined the effect of land use on road segment crashes, Karlaftis and Tarko (17), Noland and Quddus (18) and Aguero-Valverde and Jovanis (19) investigated the effects of demographic patterns and weather on county-level crashes. Some of the previously identified spatial factors, however, have not been investigated at either the corridor level or smaller scale. In particular, these spatial factors have the potential to influence crashes at intersections—the subject of this current research.

As mentioned previously, the majority of explanatory factors considered in prior studies consist of intersection traffic volume or AADT (10, 20-21) and geometric attributes of intersections (9, 22-23). Bauer and Harwood (22) indicate that traffic volume variables capture 16% to 38% of the variability in crashes, leaving a small (5 to 14%) portion of the variability explained by geometric design variables. In another study, Greibe (23) developed prediction models for both road links and junctions and was able to capture more than 60% of the systematic variation in road-level models, while intersection models had lower explanatory power. Chin and Quddus (9) confirmed like others that traffic volumes are the most important factor, or main effect, predicting crashes. While these results re-emphasize the importance of traffic volume for predicting crashes, findings by Abdel-Aty et al. (24) showed that traffic volume on the major-roadway were the most important factors only for right-turn crash prediction. In addition they mentioned that volume was not significant in all the other models they developed for various crash types. As a plausible reason for this finding, they mentioned that other variables often related to traffic volume were found to cause a larger reduction in deviance. Oh et al. (25) also found that posted speed limit and the existence of commercial driveways play an influential role for total crashes in addition to angle, left-turn, head-on, and rear-end crashes for rural intersections.

If important spatial explanatory variables are omitted from a statistical crash model then several consequences arise. First, their omission may lead to over- or under-estimation of the effect of included variables. In case of ordinary least square (OLS) estimates, secondly, parameter estimates are biased (due to lack of independence between the error term and independent variables). In the case of maximum likelihood estimate important omitted variables inflate the error term, which influences the dispersion parameter and results in incorrect variance estimates, *t*-ratios for parameter estimates, and confidence intervals. The omission of important explanatory variables also reduces the explanatory power of the model. As a result of these consequences, it is worth investigating whether spatial variables are important for explaining crashes at intersections.

The purpose of this study is three fold:

1. To show how the commonly omitted but important spatial factors can be collected in an efficient but cost effective manner;
2. To assess reasonableness of the safety effects of these spatial factors; and
3. To test the contribution of these variables in model estimation and prediction.

To achieve the last goal, two different models are developed: one with traffic volumes from major and minor-road as the only exogenous variables, and second with all the spatial variables in addition to commonly included geometric and traffic factors. The results of these two models are compared to test the significance of the spatial variables, their effect on crash occurrence, and the overall improvement in model prediction capability.

DATA DESCRIPTION AND DEVELOPMENT

In this study four different crash types, total, pedestrian, bicycle, and fatal and serious injury crashes, are analyzed. In the following two sections, the methodologies for developing the data are discussed first, followed by a detailed description of the statistical tools used to model different crash types.

Data collection and processing

Data developed and applied in this study were obtained from six different sources, including: a) crash data, b) geometric data, b) traffic volume or exposure data, c) information about traffic control parameters, d) spatial characteristics, e) weather related factors, and f) demographic data. The study sites examined in the study are signalized intersections in the City of Tucson. Intersections types were signalized intersections including four-legged and T- junctions; however, the unavailability of traffic volume data restricted the sample to 291 signalized intersections. The nature of the six data sources is now described.

Crash data

The crash data for this study were obtained from the Accident Location Identification Surveillance System (ALISS) database maintained since 1975 by the Arizona Department of Transportation (ADOT). The ALISS database contains all of the micro-level information about crashes, such as the type of crash, severity, time of occurrence, crash location and description of site, vehicle maneuvers before crash, direction of movement of the vehicle prior to the crash, information about the people involved in the crash (both driver and passenger information), as well as vehicle information. For this study data on crashes that occurred from 2001 to 2004 at 291 signalized intersections of City of Tucson were collected and analyzed. Crashes were categorized as intersection-related crashes if they occurred within the curb-line limits of the intersection or if they occurred within the influence area of the intersection, defined as within 250 ft along any leg of the intersection (from the intersection center point). A summary of various crash types is shown in Table 1. Only five intersections among the 291 intersections recorded zero total accidents over the four year period.

Geometric data

Geometric data for the signalized intersections were obtained from a City of Tucson maintained database. This database provides information about the names of the cross roads of the intersections, the unique intersection IDs, and the direction of major-roads, which helps to identify the major and minor approach of any intersection. It also contains information such as number of lanes, presence, type, and width of the left-turn bays, widths of the through lanes, and width of the right-turn bay, width and type of median, the approach speed, and the approach grade. Finally, it contains information about pedestrian crossings, left-turn bay, presence of reversible lane, and unusual circumstances like one-way streets. Based on prior research and reasoned judgment the total of list of variables was reduced to a set of explanatory variables for modeling. This list of variables as well as their summary statistics is shown in Table 2.

Traffic data

The only traffic control data that were available from the city is the number of signal phases at the intersections. Among traffic regulatory factors, speed limits of major and minor-roads were obtained. Acquiring traffic volume data was another major challenge in this study. While an ideal way to model crash data is to obtain crash and traffic volume data by year, it proved difficult to obtain annual traffic volume data for more than one year for many of the sites. Practically this means that some hypothetical site 1 might have traffic volume counts for 2001 but not for 2002 or 2003, whereas site 2 might have counts for 2002 but not for 2001 and 2003. As a consequence, calculations based on one year of traffic volume data were applied for the analysis, with all acknowledged limitations of this approach. The summary statistics of traffic control and volume related variables are also shown in Table 2.

Spatial variables

The following spatial variables represent commonly omitted factors in crash models. Some of the spatial data are related to weather (e.g. sun glare), while others are related to human factors (e.g. drunk driving). It should be pointed out that these variables span the range of what might be argued as truly causal (e.g. sun glare blocking vision which results in a crash) to surrogates for causal variables (e.g. proximity to drinking establishments which sell alcohol to individuals who may become inebriated and drive a motor vehicle). It is not claimed, therefore, that the spatial variables are perfect measures for the underlying causal mechanisms, but it is believed that all variables are capturing the effect of underlying causal mechanisms even if not directly causal.

Sun glare

Among the different types of spatial data used in this study, capturing the effect of sun-glare was a most challenging task. Andrey et al. (26) mentioned that Sun glare is a potential environmental factor having strong effect on driving performance, however, only Flahaut (27) has examined the effect of sun glare on crash occurrence. Sun glare is most problematic during early or late hours of the day, presumably within an hour after sunrise and before sunset, when the sun is on the immediate horizon. As one might expect, weather, trees, and hills will play a role on the effect of sun glare, as will the direction of travel. Also, time of year plays an important role in the intensity of glare from sunlight and the ‘critical’ hour in which glare is a potential issue. In Tucson, AZ, sun glare is especially bad in early fall and early spring when the sun rises almost exactly east and sets almost exactly west, because the Tucson road network is on a N-S E-W grid systems, is relatively flat, and is not cloudy about 350 days per year. Given these complexities it is not surprising that the potential effect of sun glare was difficult to capture in a variable.

Fortunately, mathematical expressions are available to estimate the intensity of glare from any light source. Without going into great detail, glare which is measured as equivalent veil luminance depends on the angle of glare, the angle made by the light source, and the direction of vision. In the case of sun glare, this angle is nothing but the angle of sun as seen by the motorist. As the sun’s position relative to an observer depends on the latitude and the longitude of the observer’s position on earth and time of the year, the effect of sun glare varies by time of day across months throughout the year. Hence, obtain the ‘critical’ times for potential glare, a varying ‘glare window’ needed to be determined and linked with the time and day of each crash.

The National Oceanic and Atmospheric Association (NOAA) provides sunrise and sunset times over 365 days of the year for various cities in the USA. From this source, the sunrise and sunset times were obtained for twelve months and an hour ‘windows’ immediately after sunrise and before sunset are calculated and shown in Table 3. In attempt to capture the effect of sun glare, total numbers of crashes during these specified intervals are calculated separately. Then for each intersection two different entries were considered: 1) total crashes during the possible non-glare time period i.e. over 22 hours of time per day for the 4 year period, and 2) crashes during the two hours; i.e. the morning and the evening glare period per day for 4 year period. Also, indicator variables were created to indicate if crashes were observed during glare or non-glare periods for each intersection. It is important to note that sun-glare can only occur on an eastbound approach at sunrise and a westbound approach at sunset—and so travel direction was taken into account. Finally, an offset variable is used to take the unequal period of observation, i.e. 22 hours and 2 hours time period into account. With this methodological approach it is important to note that the time of observation of crashes is used as a measure of exposure in this study. However, a theoretically better approach is to measure or estimate the traffic volumes during glare and non-glare times as predictors of glare and non-glare crashes. The absence of information about traffic volume during these two time periods restricted the ability to adopt this approach, while the exploratory nature of the research motivated the research team to pursue this line of inquiry.

School location

Two additional spatial factors examined in this study are the effect of school zones and the proximity effect of drinking locations such as bars and pubs. To process school-zone related data, first the GIS layer containing all types of schools in Arizona was obtained. This layer contains the geographic location of schools along with other attributes such as the type of school, name of school, and addresses. From this layer several new layers are created based on the school types including elementary, middle, and high schools as well as colleges and universities. Each of these layers is then mapped to the intersection GIS layers to develop indicator variables for intersections to denote proximity: intersections that fall within ¼ mile, ½ mile and 1 mile radius of each school type—resulting in 3 (radii) x 4 (school types) = 12 indicator variables. This search was performed within the GIS platform using “nearest-neighbor” analysis algorithm. This algorithm ‘looks’ for the presence of specific types of schools within a specified search radius around each intersection and creates tables showing the distance of the schools from each intersection.. From these tables produced by the search the analyst can develop the necessary indicator variables. The summary statistics of these variables are shown in Table 4.

Location of drinking establishments

While the locations of schools are obtained from the GIS layer, finding the locations of drinking establishments was not as straightforward. First of all, the availability of a GIS layer representing all the drinking location did not exist. Secondly, GIS tiger files are available that show the locations of establishments with liquor licenses; however, these may not be locations where people go, spend time and drink alcohol. For example, many supermarkets obtain liquor licenses, however it is unlikely that alcohol purchased at these locations will also be consumed there. To deal with this problem, addresses of bars and pubs in Tucson were identified from the yellow pages. Then geo-coding service in GIS was used to locate those addresses on an Arizona street map and a new layer was created showing the location of bars and pubs in Tucson (as a point). This layer was then used like the school layer to develop a 'proximity to bars' variable for bars and pubs within ¼ mile, ½ mile, 1 mile and 5 miles of the intersections. In case of bars and pubs, a higher search radius of 5 mile was considered based on the assumption that drunk drivers might drive some distance before they become involved in a crash and it is not known *a priori* whether the effects are localized or randomly spread over the transportation network. Also, drinking locations are generally found in clusters and it is not known whether the number of bars near a location or just the presence of bars affect safety. Consequently, two types of variables for each search radius were created; 1) total number of bars and pubs within a search radius, and 2) the presence or absence of drinking locations within a search radius. The summary statistics of these variables are shown in Table 5.

Weather data

To take the weather effects into account, weather data at various weather stations in the State of Arizona were obtained from National Oceanic and Atmospheric Association (NOAA). While the database provides information about hours of sunlight, cloudiness, temperature, precipitation and snow fall, information about precipitation was only available for most of the weather stations and for this reason precipitation related variables only were used. The raw database contained daily precipitation at different weather stations for all days of the month and for 12 months of the year along with the recording time of the precipitation. However, some weather stations had recorded data for less than 12 months of a year, thus only the available month's data were considered in the analysis. Two different precipitation variables were created: *Total average precipitation per year* and *Average number of rainy days per year*.

After weather stations were located using their latitude and longitude within GIS platform, a "nearest-neighbor" analysis was performed. However, large search radii were chosen because relatively few weather stations were available within the analysis region. Radii were selected such that all of the intersections were assigned weather attributes. A summary of weather related variables are shown in Table 6.

Demographic data

Finally, the demographic distribution of the location was taken into account by census track population distribution data. The census database contains distribution of population in a region and this is available in GIS layers. This GIS layer is intersected with intersection locations to identify the distribution of population around intersections—indicating of course the local driving population. A summary of demographic data are shown in Table 7.

STATISTICAL MODEL SELECTION AND DEVELOPMENT

The Poisson and NB cross-sectional models are used as the starting point of crash data modeling in this study. Panel models are then introduced to capture unobserved intersection-specific effects that influence outcomes. Fixed effects models—the usual type applied to modeling intersections—are suitable to identify the effect of any "change" in independent variables that affect the dependent variable. If an analyst was interested in unobserved intersection-specific effects, then indicator variables would be added to the models for each intersection. However, one might prefer to believe that variation among intersections is more of a random phenomenon (not fixed). In this case random effects models are appropriate. Hence in this study the random effects negative binomial (RENB) models are applied. Additional background on the Poisson, NB and RENB models is provided in Washington et al. (28) and Cameron and Trivedi (29).

As mentioned previously, one of the aims of this research is to identify the significance of the spatial variables, and also to assess their impact on safety and their contribution to explanatory power. As mentioned by Cameron and Trivedi (29), the general consequences of measurement errors in a non-linear count model are very similar to the heterogeneity that results in over-dispersion. Measurement errors in regressors can arise in various ways. For example, there could be multiplicative or additive errors in the measurement of exogenous variables as well as errors due to omission of relevant and important covariates. In this study the focus is the second case, i.e. error due to omission of relevant variables. In the case of a linear model with ordinary least square (OLS) estimated parameters, omitted variables cause bias in coefficient estimates of the included variables and it can be shown that the estimator is not consistent and results in bias estimates of coefficients. In the case of non-linear count models, the omission of relevant variables can be interpreted as unobserved heterogeneity *only* in cases where the omitted regressors are uncorrelated with the included covariates. Suppose, the included covariates are X and the omitted covariates are Z; and β and γ are the vector of parameters associated with X and Z respectively. Then the expected number of crashes can be written as:

$$\mu_i \mid X_i, Z_i = \exp(X_i^{'}\beta + Z_i^{'}\gamma) = \exp(X_i^{'}\beta)\exp(Z_i^{'}\gamma) = \exp(X_i^{'}\beta)u_i \tag{1}$$

Here $u_i = \exp(Z_i^{'}\gamma) = \exp(\varepsilon_i)$ is algebraically similar to the error component or the gamma heterogeneity term included while developing the NB model. Hence, the consequences of dropping important variables are essentially the same as those due to unobserved heterogeneity, i.e. over-dispersion and loss of efficiency of the pseudo maximum likelihood estimator.

To test the significance and importance of these commonly omitted spatial variables, two different model specifications are tested for each type of crash. In the first model, only traffic volume related variables are included, while in the second model traffic volumes as well as geometric and spatial variables are included. For example, if X is a vector of commonly included variables, and Z is a vector of commonly omitted spatial factors, then two different mean functions are tested:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots\dots\dots\dots\dots + \beta_k X_{ik} \tag{2}$$

and

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots\dots\dots + \beta_k X_{ik} + \gamma_1 Z_{i1} + \gamma_2 Z_{i2} + \dots\dots + \gamma_k Z_{ik} \tag{3}$$

where β's are vector of covariates associated with covariate vector X and γ is a vector of covariates associated with Z. To test the significance of omitted variables Z, the hypotheses are:

$$H_0 : \gamma = 0 \text{ and } H_1 : \gamma \neq 0 \tag{4}$$

The significance of the coefficients is tested using a t-test, where the test statistic is

$$t = \gamma' / se(\gamma') \tag{5}$$

which is student *t*-distributed. Recall that γ' is a vector of coefficient estimates of γ from equation 3 and $se(\gamma')$ is the vector of estimated standard errors of the coefficients. A significance level of 0.10 was used for the *t*-test for this study.

Another way to obtain the significance of coefficient is to conduct likelihood ratio test. The test statistic is given as:

$$\chi^2 \sim -2\ln\lambda$$

where $\lambda = LR/LU$, LR is the restricted likelihood from equation 2, and LU is the unrestricted likelihood from equation 3. A significance level of 0.10 was also taken for the χ^2-test for this study. For more detail on any of the test statistics the reader should refer to Washington et al. (28).

If these test results indicate that the omitted variables are significant and the model with omitted variables fit the data better, then an estimate of the change in coefficient estimates for β's from equation 2 and 3 is conducted. This change is calculated as a percentage change in estimates, with a higher estimate of β from equation 2 indicating a positive bias, and a lower estimate of β compared to equation 3 indicating a negative bias.

To measure the overall goodness of fit, the log-likelihood ratio index was calculated. For count data models the common practice is to use a "pseudo" R^2 statistic, which is often known as log-likelihood ratio index (ρ^2**).** According to Ben-Akiva and Lerman (27), ρ^2 is given by

$$\rho^2 = 1 - \frac{L(\boldsymbol{\beta})}{L(0)} \tag{6}$$

where $l(\boldsymbol{\beta})$ is the log-likelihood value of the fitted model, and $l(0)$ is log-likelihood value of the model with constant term only. Everything else being equal, a specification with a higher log-likelihood is preferred. The lowest value of log-likelihood function corresponds to the model with constant term only and is considered the base case or naïve model. The value of ρ^2 is between 0 and 1, with better models approaching the latter. Like the R^2 statistic, it has the undesirable characteristic that it will increase whenever new variables are added to the model. To overcome this disadvantage, Ben-Akiva and Lerman (27) incorporated a correction for the number of covariates, p, yielding an adjusted log-likelihood ratio index as

$$\bar{\rho}^2 = 1 - \frac{L(\boldsymbol{\beta}) - p}{L(0)} \tag{7}$$

RESULTS

This section presents the results of the signalized intersection crash models for Tucson. To investigate the effects of spatial factors on crash occurrence and to provide a means for comparison, a total intersection crash model is initially developed. Pedestrian and bicycle-related crashes are then examined. Finally, the effects of spatial factors on fatal and severe injury crashes are examined.

All of the statistical models in this study were estimated using STATA econometric software. "Best" models were selected among several competing models, considering the usual attributes such as model goodness of fit (GOF), theoretical appeal of variables, agreement with expectation, etc. All final models are negative binomial and either fixed- or random-effects. Also, for each crash type model at least two different candidate models are developed to compare models with spatial variables to models with only traffic volume. The comparisons were carried out both in terms of overall GOF as well as coefficient estimates. This enables an assessment of the relative impacts of the spatial factors and an estimate of biases caused by omitting these variables.

Total intersection crashes

The estimation results for total crash models are presented in Tables 8 and 9. A random effects negative binomial (RENB) model is used to model these crashes. The full model with all the traffic, geometric, weather, spatial and demographic covariates revealed a total of fourteen significant variables in estimating total expected intersection crashes. Similar to previous studies and not surprisingly, traffic volumes are the most important and reliable predictor of intersection crashes occurrences and are positively associated with total crashes (p-values < 0.001 for major and minor-roads). The number of signal phases is positively associated with crashes (p-value < 0.001). While the effects of number of intersection signal phases have been investigated in previous studies (9, 22, 31) there is no clear and established relationship between safety and the number of signal phases. It is presumed like Poch and Mannering (31) and Chin and Quddus (9) that higher numbers of signal phases indicates greater complexity of traffic movements within the intersection and that a considerable proportion of intersection crashes occur during or immediately following phase changes. The presence of turn lanes was not found to be statistically significant except for left-turn lanes on major-roads (p-value=0.009). Results show that the presence of the left-turn lane on major-road is associated with a higher number of crashes at intersections, which is not surprising given the possible endogeneity of this variable (11), the complexities involved with isolating this effect, and the mixed results found in prior research. The average posted speed limits of approach roads indicate some possible interesting insights. While higher minor-road speed limits are associated with increased intersection crashes (p-value=0.003), higher major-road speed limit are negatively associated (p-value=0.033). These findings should be considered in tandem, and are a bit difficult to explain. It is likely, however, that the effects of speed are residual effects of the speed-volume relationships not completely captured in the exposure term as well as effects of different design standards of facilities.

The remainder of significant variables in the model is related to spatial, demographic or weather features near to the intersections. The presence of colleges or universities within half mile of the intersection (p-value=0.013) increases total crashes. It is not surprising that this variable has a positive effect on total crash occurrence, since colleges and universities are locations with inexperienced drivers, multiple modes of travel, and complex motor vehicle movements. The proximity of bars and pubs near an intersection are significant and increase total predicted intersection crashes. The findings reveal that the effects of bars and pubs on total crashes are either localized within a quarter mile of the intersections (*p*-value < 0.001) or very far away within 1 to 5 miles of the intersections (*p*-value < 0.001). This finding, though not conclusive, suggests that the proximity of drinking establishments nearby to intersections increases the number of expected crashes, presumably due to a greater-than-average number of intoxicated pedestrians, bicyclists, and drivers getting involved in crashes. It is possible, however, that the 'bar effect' is picking up some other land-use related unknown effect.

Similar to the previous spatial factors, the demographic patterns around the intersection are associated with total crash frequencies. Results show that intersections in regions with proportionately larger population between 0 and 15 years of age (*p*-value=0.066) have more crashes, and intersections in regions with proportionately larger population aged 16 and 64 have fewer crashes (*p*-value=0.027). The variable for the population group above 65 years of age was not significant.

Finally, the weather variables such as annual average precipitation (*p*-value=0.033) and annual average number of rainy days (*p*-value=0.021) had negative relationships with crash occurrence. Rain in Tucson is quite rare, with about 12 inches rainfall per year on average. Thus, it is believed that drivers may exercise greater caution on rainy days than during other times.

The dummy variable for the glare effect (*p*-value < < 0.0011) is positively correlated with crash occurrence. The finding suggests that the presence of sun glare increases traffic crashes—obstructing the normal visibility to drivers after sunrise and before sunset when the sun is almost at the horizon. However, the limitation in how this variable is coded suggests that this variable may partially be capturing the effect of congestion, since sunrise and sunset during certain times of the year coincide with the peak period (see Table 3).

A comparison with the total crash model with traffic volumes as the sole predictors show a poor overall fit of about $\rho^2 = 0.01$, whereas the model with the spatial variables fitted the data better with ρ^2 is 0.719. This indicates that the spatial variables themselves explain a large portion of the variability. To investigate the impact of the spatial variables on crash occurrence, the marginal effects of all the significant variables are calculated. The comparisons of the marginal effects reveal that the dummy variable for glare has the largest elasticity. As mentioned previously, it is likely that the glare-related dummy variable captured additional temporal and spatial effects, mainly the effects of peak hour traffic, when the hourly traffic volume is significantly higher than average daily traffic volume. With the increase in volume potential for conflicts also increases. Hence, the results from spatial variable model suggest that traffic volume is an important factor in explaining crash occurrence, but emphasizes the importance of spatial and temporal covariates in crash modeling. The comparisons of coefficients estimates also show that the effect of major-road traffic volume is biased downwards whereas the effect of minor-road traffic volume is biased upwards if spatial factors are excluded from model building.

Pedestrian and bicycle crashes

Pedestrians and bicyclists are vulnerable or 'weak' road users that warrant investigation. An initial investigation showed that sun glare had no effect on these types of crashes—hence the glare effect is omitted from consideration in these models. NB cross-sectional count data models were developed for each of these crash models, the outcomes of which are shown in Tables 10 and 11.

Seven variables were found to be statistically associated with pedestrian-involved crashes. Like the total crash model, AADT on major and minor-roads were positively associated with pedestrian crashes (*p*-value= 0.056 and 0.002 for major-road and minor-road AADT respectively). Of course, pedestrian crashes depend on pedestrian volumes and an intersection with high pedestrian and traffic interaction would reveal a different safety experience than an intersection with low pedestrian volume. However, as is common, pedestrian exposure was not available for this study and thus the usual limitations of traffic exposure apply. The number of signal phases is positively associated with pedestrian crashes (*p*-value= 0.018). This result is likely indicative of intersections with higher complexity and higher pedestrian exposure. No geometric design characteristics were found to influence pedestrian crashes.

Besides the traffic-related covariates, three spatial and weather-related factors were significant. Among the spatial factors, the presence of bars and pubs within a quarter mile of the intersections was very significant (*p*-value < 0.001). This result suggests that the effect of the drinking establishments on pedestrian crashes is localized within the quarter mile radius of their locations—a distance often cited as the extent of 'normal' walking distances. Also, the presence of elementary schools within one mile of an intersection, and colleges and universities within a half mile of an intersection were found to be positively associated with pedestrian crashes (*p*-value= 0.093 and 0.165 for elementary and colleges and universities respectively). Due to high pedestrian activities near elementary schools, colleges and universities, an increase in pedestrian crashes at intersections is expected. Finally, the annual average precipitation is negatively correlated (*p*-value= 0.007) with crash occurrence. This is intuitive as pedestrians are less likely to walk on rainy days. The ρ^2 of this model is 0.094, which indicates that about 10% of the variability in pedestrian crashes is captured by the explanatory variables in the model. The unexplained portion of this model is considerably higher than the total crash model and is likely to be improved by incorporating information regarding pedestrian activity and volume.

The results from the negative binomial bicycle crash models are presented in Tables 12 and 13. It is no surprise that major and minor-road AADT are significant (*p-value* < 0.001) in explaining bike crashes at intersections. With higher traffic exposure, involvement of bike-involved crashes increases. Similar to pedestrian crashes, bicycling exposure would provide an improved measure for prediction. No geometric design elements of intersections were found to influence bike crashes at intersections. While the presence of bars and pubs within a half mile of the intersections is positively associated with bike crashes (*p*-value = 0.006), the presence of bars and pubs within a quarter mile of the intersections is only slightly increased (*p*-value = 0.017). The combined results of these two findings needs to be considered—since a bicycle crash occurring within a half mile of an intersection gets a ‘correction’ for both within a quarter mile and within a half mile—with the net effect being an increase in crashes near to bars and pubs but decreasing with increasing distance. Similar to total and pedestrian crashes, the presence of colleges and universities within a half mile of the intersections has positive effect on bicycle-involved crashes (*p*-value= 0.035). Colleges and universities are locations with increased bicycle usage by relatively inexperienced road users. As a result, these locations are consistently identified to be sensitive to traffic accidents. Also, the presence of middle schools within 1 mile of intersections is significant (*p*-value=0.016) and positively associated with bike crashes. The prevalence of young bike riders near middle schools clearly explains this result.

Fatal and severe injury crashes

Fatal and incapacitating injury crashes were examined to assess the effects of the spatial factors on these crashes. Sun glare does not appear to be important for explaining these crashes. A NB model was developed and the results are shown in Tables 14 and 15. The overall goodness-of-fit statistics showed a ρ^2 value of 0.099. Traffic volume on major and minor-roads are positively correlated with injury crashes (*p-value* < 0.001 for both major-road and minor-road volume). Oh et al. (25) also showed that higher major and minor-road traffic volumes resulted in higher injury crashes at rural signalized intersections in Georgia. Similar to total crashes, higher numbers of signal phases were found to increase fatal and injury crashes (*p*-value=0.026). Also, left-turning lanes (*p*-value=0.009) on the major-road increases severe injury crashes, while the presence of right-turn lanes on major-road (*p*-value=0.018) is associated with fewer crashes. Posted speed limits on the minor-road (*p-value* < 0.001) revealed similar effects as total crash models and are correlated with higher number of severe intersection crashes. The presence of a negative gradient on the major-road is positively correlated (*p*-value=0.034) with severe injury crashes. Among the four models developed in this study, this variable was significant only in this model. Downward gradients increase vehicle speeds and increase stopping sight distances. While the spatial variables related to location of schools as well as bars and pubs were significant in all the other models, these variables were not significant in this model. This observation indicates that spatial variables have no significant effect on fatal and injury crashes at intersections—however one might suspect that spatial variables might play a role on high speed roads. In addition, annual average precipitation and annual average number of rainy days and sun glare did not have any effect on severe injury crashes. However, demographic patterns near to intersections is important, with the population of age between 0 and 15 positively (*p-value* <0.001) and the population between age between 16 and 64 negatively associated with intersection-related fatal and injury crashes.

DISCUSSION OF FINDINGS

Discussed here are the elasticities of various factors in the models and the four models are compared and contrasted (see Table 16). An elasticity is an estimate of the percentage change in the dependent variable due to a 1% change in the independent variable, with all other variables computed at their means. The elasticity provides a measure of the relative influence of various independent variables on intersection safety.

In agreement with numerous previous studies, traffic volume is an influential predictor of crash occurrence at intersections. Results from this study also showed that all four types of crashes increase with increase in the log of major and minor-road traffic volume. Elasticity for the major-road AADT ranged from 3.6 to 7.13, with the highest effect on fatal and injury crashes and that of minor-road AADT ranged between 1.9 and 2.8. The difference in elasticity range gives an idea about the relative impact of major and minor-road traffic volume on safety. The elasticity of 7.13 for fatal and injury crashes indicates that a 1% increase in the log of major-road traffic volume leads to about a 7% increase in fatal and severe crashes for a 4-year time period. The influence of major-road traffic volume is smallest in the case of pedestrian crashes. Minor-road traffic volumes reveal the highest elasticity for pedestrian crashes. This finding suggests that intersections with relatively high traffic volume on minor-roads are associated with greater pedestrian activity and careful attention to pedestrian safety should be provided at those intersections. Similarly, major-road traffic has a significant impact on bicycle-related crashes.

The number of signal phases is found to be positively correlated with the total, pedestrian, and fatal and injury crashes. As identified previously, higher numbers of signal phases are common in busy intersections and this is probably a surrogate for complex intersections and activity at those intersections. The elasticity however shows that signal phase has a large impact on pedestrian crash that shows 1% increase in the number of phases increases pedestrian crash by 0.74% (so a 100% increase from 2 to 4 phases would increase crashes by 74%). Whereas the percentage increase for total and fatal crashes are about 0.38 and 0.36 respectively.

Among geometric design elements, the presence of turning lanes revealed a significant influence on safety. As mentioned previously, many prior studies observed mixed effects of left-turn lanes on safety, in part due to the potential endogeneity of this variable. The results indicate that the presence of left-turn lanes increases fatal crashes by 1.8 times and total crashes by 1.6 times. The presence of a right-turn lane was found to reduce fatal crashes by 0.78 times, suggesting that separating slow-moving and right-turning traffic from the through traffic significantly improves intersection safety.

As mentioned by Oh et al. (25), the role of speed is a primary concern in safety, but the effect of posted speed on safety is murky. Some studies have indicated that faster approach speeds are not as hazardous and perhaps safer because turning and lane changing maneuvers occur more frequently when the posted speeds are lower. However, other studies revealed that higher speeds are associated with increased crash involvement and severities as speeding increases stopping sight distance. The results from this study indicate that higher major-road posted speeds are associated with fewer total crashes, and more total crashes on minor roads with higher speeds. It is believed that posted speeds are poor measures of operating speeds and capture to a large extent the design standards of roads, and thus are confounded in this research.

Only the fatal crash model showed that downward gradient is associated with increased crashes. The elasticity indicates that the presence of downward gradient on major-roads increases expected fatal crashes by 1.5 times. Negative gradients increases stopping distance and can sometimes reduce visibility of an upcoming intersection.

While the impact of geometric and traffic factors are somewhat known through prior traffic safety research, the impact of spatial factors is relatively unexamined. This study examined a variety of spatial factors that have heretofore not been the subject of safety research. It should be noted that while solid theory supports the spatial effects being examined, ecological fallacy and confounding may be playing a role in the modeling results.

This study incorporated a total of 12 variables related to the spatial location of elementary, middle, and high schools as well as colleges and universities. The presence of colleges and universities within a half mile of intersections was found to increase total, pedestrian, and bike crashes by 31%, 39% and 45% respectively. Pedestrian crashes at intersections within 1 mile of elementary schools increase by a factor of 1.6 and bike crashes at intersections within 1 mile of middle schools are increased by a factor of 0.51 on average. These results suggest that the increased pedestrian and bicycle activity near these facilities is an important factor at nearby intersections.

While blood alcohol concentration (BAC) is known to influence traffic safety, this study investigated the spatial correlation of nearby drinking establishments with various intersection crashes. Intersections with bars and pubs within a quarter mile are expected to observe about 1.3 times more total crashes and 1.97 times more pedestrian crashes than otherwise similar intersections. In addition, it is also observed that bike crashes would be affected by as much as 1.6 times in the presence of bars and pubs within half mile of the intersections. These findings are intuitive in that frequently cited normal walking distance is ¼ mile and bicycle distances should be greater. It should be noted, however, that the spatial scale of these variables might be confounded with other affects, like urban effects (tall buildings, complex intersections, many modes of travel, etc.), and thus may suffer from ecological correlation.

Inspection of the effect of precipitation on intersection safety indicated that total as well as pedestrian crashes are less with increases in both annual average precipitation and annual average number of rainy days. While precipitation also has revealed mixed effects on safety as identified by Shankar et al. (32), Zang and Holm (33), the results are intuitive as the data were obtained from Tucson, Arizona. While travel demand is not affected in states with higher average precipitation and higher number of rainy days, demand in places like Tucson can be greatly influenced by occasional rain. Also, drivers lack experience driving in the rain and may be more cautious, especially at intersections.

The research found a positive relationship between sun glare and traffic crashes. The percentage of crashes during possible glare time is disproportionately high, indicating that the presence of the effect of sun glare increases total crashes by a factor of 1.2. Including this variable in the model significantly increased model explanatory power

(40%). Because this factor may also be picking up congestion effects, a recommended improvement is to parse exposure by glare and non-glare periods—a data break down that could not be performed for this analysis.

CONCLUSIONS AND RECOMMENDATIONS

This research effort explores crash occurrence at intersections by including typically excluded contributory factors. The examination of the typically omitted spatial factors across various crash types clearly demonstrates both statistically and through reasoned logic, that spatial factors may exert a significant influence on intersection safety. While some prior research has examined the effect of spatial factors at county or census-track level models, this is first known effort to incorporate spatial variables in intersection crash models. It is emphasized that confounding and ecological correlation are possible at some of the spatial scales examined in this paper, and so conclusions are meant to serve as entry points for further inquiry rather than representing the final word on these spatial factors. Based on the results of this research the following conclusions are drawn:

1) Special traffic generators, specifically schools and universities, result in increased pedestrian and bicycle activities near these locations. Combined with the inexperience of these road users, involvement in pedestrian, bicycle, and motor-vehicle intersection crashes near these special traffic generators is significantly increased.
2) Drunk driving is a serious problem in the US, and contributes to many injury and fatal crashes. Drunk drivers must be undertaking a trip to pose a risk, which often originates from a drinking establishment (to home or other location). Not surprisingly, intersections near clusters of drinking establishments experience higher numbers of crashes.
3) Periods of extreme sun glare can drastically reduce the ability to safely operate a motor vehicle. Some drivers routinely commute during periods of extreme sun glare. Some road networks are more susceptible to sun glare related problems, particularly eastbound morning travel and westbound afternoon travel during periods of high glare. It should not be surprising that roadways carrying a significant portion of traffic during these times would observe greater numbers of crashes.
4) The importance of omitting spatial factors cannot be overstated. Many aspects of roadway safety management may suffer from omitting these variables. Perhaps most importantly, hot spots might be identified for geometric or other improvements when in fact crash counts are elevated due to spatial effects. An engineer might improve the intersection without realizing benefits, since the elevated crash counts are unrelated to geometric features. In modeling, only correlated variables (with spatial variables) included in a model can 'explain' the effects of the omitted spatial variables—thus over- or under-stating the effect of included variables. In short, omitting these variables will lead to inaccurate estimates of safety.

Based on these conclusions, some recommendations are also offered:

1) As the collective understanding of road safety is to improve the existing knowledge about safety, the quality and quantity of data must be improved along with improvements in analytical methods. Spatial factors appear to influence safety, are intuitive, and point to the need to broaden our collective view of factors that should be considered in safety prediction.
2) Spatial factors also provide a way to introduce behavioral factors into safety prediction and understanding. Many of the spatial factors considered here imply behavioral impacts to the transportation system.
3) While much was learned through this exercise, improvements for future research are possible and necessary. Sun-glare exposure should reflect volumes during times of sun glare and during non-glare times. Also, DUI crashes should be culled from the data and examined in relation to bars, in addition to total, pedestrian, and bicycle crashes. A relationship of bar and pub proximity with DUI crashes would strengthen the belief that this spatial factor is indeed capturing the effect of intoxicated drivers. Unfortunately, reporting rates of alcohol and drug use may prove this improved analysis difficult.

REFERENCES

1. Joshua, S. C., and N. J.Garber. Estimating Truck Accident Rate and Involvements Using Linear and Poisson Regression Models, *Transportation Planning and Technology*, Vol. 15, No.1, 1990, pp.41-58.

2. Miaou, S., and H. Lum. A Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. *Transportation Research Record 1407,* TRB, National Research Council, Washington, D.C., 1993, pp. 11-23.

3. Miaou, S., The relationship Between Truck Accidents and Geometric Design of Road Section: Poisson Versus Negative Binomial Regression, *Accident Analysis and Prevention*, Vol. 26, No. 4, 1994, pp. 471-482.

4. Shankar, V., J. Milton., and F. Mannering. Modeling Accident Frequencies as Zero-Altered Probability Process: An Empirical Inquiry. *Accident Analysis and Prevention*, Vol. 29, No. 6, 1997, pp. 829-837.

5. Mitra, S., H.C. Chin, and M. A. Quddus. Study of intersection accidents by maneuver type, *Transportation Research Record 1784*, 2002, pp. 43 – 50.

6. Kumara, S. P. and H.C. Chin. Modeling Accident Occurrence at Signalized Intersections with Special Emphasis on Excess Zeros, *Traffic Injury Prevention*, Vol. 4, No. 1, 2003, pp. 53-57.

7. Lord, D., S. Washington, and J. Ivan. Poisson, Poisson-Gamma, and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention*, Vol. 37, No. 1, 2004, pp 35-46.

8. Shankar, V., A. Richard., J. Milton., and F. Mannering. Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model. *Transportation Research Record 1635*, TRB, National Research Council, Washington, D.C., 1998, pp. 44-48.

9. Chin, H. C. and M. A. Quddus. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis and Prevention*, Vol. 35, 2003, pp. 253 – 259.

10. Kim, Do-Gyeong, Y. Lee, S. Washington, and K. Choi. “Modeling Crash Outcome Probabilities at Rural Intersections: Application of Hierarchical Binomial Logistic Models”. Accident Analysis & Prevention, In Press. 2006.

11. Kim, Do-Gyeong and S. Washington. The Significance of Endogeneity Problems in Crash Models: An Examination of Left-Turn Lanes in Intersection Crash Models. Submitted to Accident Analysis and Prevention, In Press. 2006.

12. Hauer, E., J. C. Ng, and J. Lovell. Estimation of safety at signalized intersections. *Transportation Research Record 1185*, 1988, pp. 48-61.

13. Maher, J. and I. Summersgill. A Comprehensive methodology for the fitting of predictive accident models. *Accident Analysis and Prevention* Vol. 28, No. 3, 1996, 281-296.

14. Lord, D. and B. N. Persaud. Accident prediction models with and without trend: Application of the generalized estimating equations procedure. *Transportation Research Record 1717*, 2000, pp. 102-108.

15. Ivan, J. N., C. Wang, and N. R. Bernardo. Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis and Prevention*, Vol. 32, No. 6, 2000, 787-795.

16. Ossenbruggen, P. J., J. Pendherker. and J. N. Ivan. Roadway safety in rural and small urbanized areas. *Accident Analysis and Prevention*, Vol. 33, No. 4, 2001, pp. 485-498.

17. Karlaftis, M. G., and Tarko, A. P. 1998. Heterogeneity considerations in accident modeling. *Accident Analysis and Prevention,* Vol. 30, No. 4, pp. 425-433.

18. Noland, R. B., and Quddus, M. A. 2004. A spatially disaggregate analysis of road casualties in England. *Accident Analysis and Prevention*, Vol. 36, No. 6, pp. 973-984.

19. Aguero-Valverde, J., and Jovanis, P. P. 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis and Prevention*, Vol. 38, No. 3, pp. 618-625.

20. Persaud, B., and T. Nguyen. Disaggregate safety performance models for signalized intersections on Ontario provincial roads. *Transportation Research Record 1635*, 1998, pp. 113-120.

21. Lord, D., and Persaud, B. N. 2000. Accident prediction models with and without trend: Application of the generalized estimating equations procedure. *Transportation Research Record 1717*, pp. 102-108.

22. Bauer, K. M., and Harwood, D. W. 1996. Statistical Models of At-Grade Intersection Accidents. Federal Highway Administration.

23. Griebe, P. Accident Prediction Models for Urban Roads. *Accident Analysis and Prevention*, Vol. 35, No. 2, pp. 273-285.

24. Abdel-Aty, M., J. Keller, and P. A. Brady. Analysis of the types of crashes at Signalized Intersections using complete crash data and tree-based regression. *Annual meeting of Transportation Research Board*, Washington. D.C. 2005.

25. Oh, J., S. P. Washington, and K. Choi. Development of accident prediction models for rural highway intersections. *Transportation Research Record 1897*, 2004, pp. 18-27.

26. Andrey, J., B. Mills, and J. Vandermolen, Weather Information and Road Safety, *Weather and Transportation in Canada*, 2001.

27. Flahaut, B. Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation. *Accident Analysis and Prevention*, Vol. 36, No.6, 2004, pp. 1055-1066.

28. Washington, S., M. Karlaftis, and F. Mannering. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall, Boca Raton, 2003.

29. Cameron, A. C. and P. K. Trivedi, Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK, 1998.

30. Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Massachusetts, 1985.

31. Poch, M. and F. Mannering. Negative Binomial Analysis of Intersection Accident Frequencies. Journal of Transportation Engineering, Vol. 122, No. 2, 1996, pp. 105-113.

32. Shankar, V. N., F. Mannering., and W. Barfield. Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention*, Vol. 27, No. 3, 1995, pp. 371-389.

33. Zhang, L., and Holm, P. Identifying and Assessing Key Weather-Related Parameters and their Impacts on Traffic Operations using Simulation. Federal Highway Administration Turner Fairbanks Highway Research Center, 2004.

LIST OF TABLES

TABLE 1 Summary statistics of intersection crashes in Tucson from 2001-2004

Variable Abbreviation	Variable Description	Mean	Std. Deviation	Maximum	Minimum
TotCrash	Total intersection crash	53.305	44.449	246	0
PedCrash	Total pedestrian crash	0.896	1.518	14	0
SevInjCr	Total fatal and serious injury crash	2.567	2.662	15	0
BikeCrsh	Total bike related crash	0.938	1.271	9	0
SinVCrsh	Total single vehicle crash	3.749	3.438	26	0
SidSwpCr	Total side swipe crash	3.694	4.315	34	0
AngCrsh	Total angle crash	8.714	6.670	36	0
LftrnCr	Total left-turn crash	12.432	13.009	74	0
RrendCr	Total rear-end crash	22.261	22.059	138	0

TABLE 2 Summary statistics for geometric and traffic variables

Variable Name	Variable Description	Mean	Std. Deviation	Maximum	Minimum
Phase	Number of signal phases at the intersection	2.825	0.875	4.00	1.00
ADTMAJ	Average daily traffic from major-road	31511.892	13457.927	66364.29	1617.86
LNADTMJ	log of average daily traffic from major-road	10.236	0.559	11.10	7.39
ADTMIN	Average daily traffic from minor-road	13704.486	11275.519	49296.43	0.00
LNADTMN	log of average daily traffic from minor-road	9.105	1.194	10.81	0.00
LFTMAJ	Presence of left-turn lane in major direction (1 if present, otherwise 0)	0.945	0.228	1.00	0.00
LFTMIN	presence of left-turn lane in minor direction (1 if present, otherwise 0)	0.911	0.286	1.00	0.00
RTMAJ	presence of right-turn lane in major direction (1 if present, otherwise 0)	0.340	0.475	1.00	0.00
RTMIN	presence of right-turn lane in minor direction (1 if present, otherwise 0)	0.522	0.500	1.00	0.00
MediaWd	Width of median (ft)	2.476	2.864	9.43	0.00
SpdMAJ	Posted speed in major direction (mph)	37.844	5.568	55.00	25.00
SpdMIN	Posted speed in minor direction (mph)	32.062	7.299	55.00	0.00
DwnGrdMAJ	Presence of downhill grade in major direction	0.024	0.153	1.00	0.00
DwnGrdMIN	Presence of downhill grade in minor direction	0.014	0.117	1.00	0.00

TABLE 3 Typical morning and evening sun glare window

Months	Morning Window(AM)	Evening Window(PM)
Jan	7.30-8.30	4.30-5.30
Feb	7.30-8.30	5-6
March	6.45-7.45	5.30-6.30
April	6-7	5.55-6.55
May	5.30-6.30	6-7
June	5.20-6.20	6.30-7.30
July	5.25-6.25	6.40-7.40
August	5.40-6.40	6.20-7.20
September	6.10-7.10	5.45-6.45
October	6.30-7.30	5-6
November	6.50-7.50	4.30-5.30
December	7.20-8.20	4.20-5.20

TABLE 4 Summary statistics for school location variables

Variable Name	Variable Description	Mean	Std. Deviation	Maximum	Minimum
EleSc1M	Presence of elementary school within 1 mile of intersection (1 if present, 0 otherwise)	0.948	0.221	1.00	0.00
EleScHM	Presence of elementary school within half mile of intersection (1 if present, 0 otherwise)	0.584	0.494	1.00	0.00
EleScQM	Presence of elementary school within quarter mile of intersection (1 if present, 0 otherwise)	0.213	0.410	1.00	0.00
MidSc1M	Presence of middle school within 1 mile of intersection (1 if present, 0 otherwise)	0.715	0.452	1.00	0.00
MidScHM	Presence of middle school within half mile of intersection (1 if present, 0 otherwise)	0.275	0.447	1.00	0.00
MidScQM	Presence of middle school within quarter mile of intersection (1 if present, 0 otherwise)	0.072	0.259	1.00	0.00
HigSc1M	Presence of high school within 1 mile of intersection (1 if present, 0 otherwise)	0.698	0.460	1.00	0.00
HigScHM	Presence of high school within half mile of intersection (1 if present, 0 otherwise)	0.340	0.475	1.00	0.00
HigScQM	Presence of elementary school within quarter mile of intersection (1 if present, 0 otherwise)	0.162	0.369	1.00	0.00
ColUni1M	Presence of college or university within 1 mile of intersection (1 if present, 0 otherwise)	0.392	0.489	1.00	0.00
ColUniHM	Presence of college or university within half mile of intersection (1 if present, 0 otherwise)	0.134	0.341	1.00	0.00
ColUniQM	Presence of college or university within quarter mile of intersection (1 if present, 0 otherwise)	0.034	0.182	1.00	0.00

TABLE 5 Summary statistics for bars and pubs related variables

Variable Name	Variable Description	Mean	Std. Deviation	Maximum	Minimum
Pubs1M	Number of pubs within 1 mile of intersection	6.619	9.168	31.00	0.00
Pub1Mind	Presence of pubs within 1 mile of intersection (1 if present, 0 otherwise)	0.756	0.430	1.00	0.00
PubsHM	Number of pubs within half mile of intersection	2.643	5.669	23.00	0.00
PubHMInd	Presence of pubs within half mile of intersection (1 if present, 0 otherwise)	0.457	0.499	1.00	0.00
PubsQM	Number of pubs within quarter mile of intersection	0.852	2.420	23.00	0.00
PubQMInd	Presence of pubs within quarter mile of intersection (1 if present, 0 otherwise)	0.271	0.445	1.00	0.00
Pubs5M	Number of pubs within 5 mile of intersection	70.646	31.773	118.00	0.00
Pub5Mind	Presence of pubs within 5 mile of intersection (1 if present, 0 otherwise)	0.983	0.130	1.00	0.00

TABLE 6 Summary statistics for weather variables

Variable Name	Variable Description	Mean	Std. Deviation	Maximum	Minimum
AvgPrcp	Average precipitation near the intersection	902.575	56.080	1064.333	51.67
AvgRDay	Average number of rainy day near the intersection	48.061	4.411	56.080	4.41

TABLE 7 Summary statistics for demographic variables

Variable Name	Variable Description	Mean	Std. Deviation	Maximum	Minimum
POPTOT	Total population near the intersection	1256.124	1318.717	11190.00	168.00
POPURB	Total population in urban area near the intersection	1217.478	1287.602	11100.00	168.00
POP00_15	Total population from age 0 to 15 years old near the intersection	293.591	420.906	3450.00	0.00
POP16_64	Total population from age 16 to 64 years old near the intersection	821.052	859.494	7322.00	82.00
POP65	Total population over 65 years old near the intersection	141.473	119.552	618.00	5.00

TABLE 8 Results for total intersection crashes (RENB model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-13.21	-0.05	0.960
Log of AADT on the major-road	0.6169	7.81	< 0.001
Log of AADT on the minor-road	0.1827	4.38	< 0.001
Number of phases at the intersection	0.1805	4.03	< 0.001
Left-turn lane indicator (1 if at least one left-turn lane on the major-road, 0 otherwise)	0.4722	2.62	0.009
Average speed along the major direction (mph)	-0.0223	-2.14	0.033
Average speed along the minor direction (mph)	0.0230	2.99	0.003
College or university within half mile indicator (1 if at least college or university within ½ mile of the intersection, 0 otherwise)	0.2739		
Bars within quarter mile indicator (1 if at least one bar or pub within ¼ mile of the intersection, 0 otherwise)	0.3181	3.59	< 0.001
Bars within five miles indicator (1 if at least one bar or pub within 5 miles of the intersection, 0 otherwise)	2.440	3.58	< 0.001
Total population between age 0 and 15 near the intersection	0.0016	1.84	0.066
Total population between age 16 and 64 near the intersection	- 0.0013	-2.21	0.027
Average annual precipitation near intersection	-0.0017	-2.13	0.033
Average annual number of rainy days near intersection	-0.0217	-2.31	0.021
Glare indicator (1 if crash occurred during glare period, 0 otherwise)	0.1883	7.00	< 0.001
Parameter a	2876775		
Parameter b	2.98		
Number of observations	582		
Number of groups	291		
Log-likelihood at zero	-6738.01		
Log-likelihood at convergence	-1893.68		
ρ^2	0.719		
AIC	3817.36		

TABLE 9 Results for total intersection crashes with traffic volume only (RENB model)

Variables	Estimated Coefficient	*t*-statistic	*p*-value
Constant	-28.80	-39.25	< 0.001
Log of AADT on the major-road	0.5531	8.34	< 0.001
Log of AADT on the minor-road	0.3679	9.19	< 0.001
Parameter a	4.37		
Parameter b	17.41		
Number of observations	582		
Number of groups	291		
Log-likelihood at zero	-6738.00		
Log-likelihood at convergence	-6675.62		
ρ^2	0.009		
AIC	13357.24		

TABLE 10 Results for intersection pedestrian crashes (NB model)

Variables	Estimated Coefficient	*t*-statistic	*p*-value
Constant	-2.399	-0.71	0.479
Log of AADT on the major-road	0.361	1.91	0.056
Log of AADT on the minor-road	0.310	3.08	0.002
Number of phases at the intersection	0.263	2.37	0.018
College or university within half mile indicator (1 if at least college or university within ½ mile of the intersection, 0 otherwise)	0.332	1.39	0.165
Elementary school within 1 mile indicator (1 if at least college or university within 1 mile of the intersection, 0 otherwise)	0.973	1.68	0.093
Bars within quarter mile indicator (1 if at least one bar or pub within ¼ mile of the intersection, 0 otherwise)	0.683	3.76	< 0.001
Average annual precipitation near intersection	-0.007	-2.71	0.007
Number of observations	291		
Dispersion parameter, alpha	0.779		
Log-likelihood at zero	-377.793		
Log-likelihood at convergence	-343.795		
ρ^2	0.09		
AIC	703.59		

TABLE 11 Results for intersection pedestrian crashes with traffic volume only (NB model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-10.0181	-5.19	< 0.001
Log of AADT on the major-road	0.5856	3.15	0.002
Log of AADT on the minor-road	0.2537	5.07	< 0.001
Dispersion parameter alpha	0.9969		
Number of observations	291		
Log-likelihood at zero	-379.5773		
Log-likelihood at convergence	-361.2933		
ρ^2	0.048		
AIC	728.58		

TABLE 12 Results for intersection bicycle crashes (NB model)

Variables	Estimated Coefficient	*t*-statistic	*p*-value
Constant	-10.3785	-6.17	< 0.001
Log of AADT on the major-road	0.6933	4.27	< 0.001
Log of AADT on the minor-road	0.2851	3.70	< 0.001
Bars within half mile indicator	0.4734	2.73	0.006
Bars within quarter mile indicator	-0.4563	-2.38	0.017
College or university within half mile indicator (1 if at least college or university within ½ mile of the intersection, 0 otherwise)	0.3762	2.10	0.035
Middle school within 1 mile indicator (1 if at least college or university within 1 mile of the intersection, 0 otherwise)	0.4126	2.41	0.016
Dispersion parameter alpha	0.2580		
Number of observations	291		
Log-likelihood at zero	-389.1658		
Log-likelihood at convergence	-356.5841		
ρ^2	0.084		
AIC	727.16		

TABLE 13 Results for intersection bicycle crashes with traffic volume only (NB model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-10.4893	-6.08	< 0.001
Log of AADT on the major-road	0.7573	4.55	< 0.001
Log of AADT on the minor-road	0.2787	3.48	< 0.001
Dispersion parameter alpha	0.3855		
Number of observations	291		
Log-likelihood at zero	-389.1658		
Log-likelihood at convergence	-367.0641		
ρ^2	0.05		
AIC	740.12		

TABLE 14 Results for fatal and incapacitating intersection crashes (NB model)

Variables	Estimated Coefficient	*t*-statistic	*p*-value
Constant	-31.6949	-27.06	< 0.001
Log of AADT on the major-road	0.6974	6.25	< 0.001
Log of AADT on the minor-road	0.2147	2.87	0.004
Number of phases at the intersection	0.1287	2.23	0.026
Left-turn lane indicator (1 if at least one left-turn lane on the major-road, 0 otherwise)	0.7684	2.62	0.009
Right-turn lane indicator (1 if right-turn lane on the major-road, 0 otherwise)	-0.2377	-2.36	0.018
Average speed along the minor direction (mph)	0.0378	3.69	< 0.001
Downhill grade along major-road (1 if exists, 0 otherwise)	0.5323	2.13	0.034
Total population between age 0 and 15 near the intersection	0.0013	3.64	< 0.001
Total population between age 16 and 64 near the intersection	-< 0.0017	-4.13	< 0.001
Number of observations	582		
Number of groups	291		
Log-likelihood at zero	-1883.63		
Log-likelihood at convergence	-686.96		
ρ^2	0.6353		
AIC	1397.92		

TABLE 15 Results for fatal and incapacitating intersection crashes with traffic volume only (NB model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-10.4893	-6.08	< 0.001
Log of AADT on the major-road	0.7573	4.55	< 0.001
Log of AADT on the minor-road	0.2787	3.48	< 0.001
Dispersion parameter alpha	0.3855		
Number of observations	291		
Log-likelihood at zero	-389.1658		
Log-likelihood at convergence	-367.0641		
ρ^2	0.05		
AIC	740.12		

TABLE 16 Elasticity for all the four intersection crash models (NB model)

	Estimated elasticity			
Variables	Total crash	Pedestrian crash	Bike crash	Fatal & severe crash
Log of AADT on the major-road	6.1269	3.6954	7.0965	7.1386
Log of AADT on the minor-road	1.8235	2.8300	2.5955	1.9550
Number of signal phases	0.3872	0.7442	-	0.3636
Left-turn lane on major-road *	0.6036	-	-	1.156
Right-turn lane on minor-road *	-	-	-	-0.2115
Average speed along the major direction	-0.6672	-	-	-
Average speed along the minor direction	0.7439	-	-	1.2104
Downhill grade along major-road *	-	-	-	0.702
Elementary school within 1 mile *	-	1.646	-	-
Middle school within 1 mile *	-	-	0.5108	-
College or university within half mile indicator *	0.3151	0.3936	0.4568	-
Bars within quarter mile indicator *	0.3745	0.9797	-0.3663	-
Bars within half mile indicator *	-	-	0.6054	-
Bars within five miles indicator *	10.47	-	-	-
Total population between age 0 and 15 near the intersection	0.1735	-	-	0.3671
Total population between age 16 and 64 near the intersection	-0.3142	-	-	-0.6002
Average annual precipitation	-1.5005	-6.4500	-	-
Average annual number of rainy days	-1.0682	-	-	-
Dummy variable for glare *	0.2112	-	-	-

* Indicator variables